

RTGWG
Internet-Draft
Intended status: Standards Track
Expires: April 28, 2022

S. Bryant, Ed.
University of Surrey 5GIC
U. Chunduri, Ed.
Intel Corporation
A. Clemm
Futurewei
October 25, 2021

Preferred Path Routing Framework
draft-chunduri-rtgwg-preferred-path-routing-01

Abstract

Capacity demands, Traffic Engineering (TE) and determinism are some of key requirements for various cellular, edge and industrial deployments. These deployments span from many underlying data plane technologies including native IPv4, native IPv6 along with MPLS and Segment Routing (SR).

This document provides a framework for Preferred Path Routing (PPR). PPR is a method of providing path based dynamic routing for a number of packet types including IPv4, IPv6 and MPLS. This seamlessly works with a controller plane which holds both complete network view of operator policies, while distributed control plane providing self-healing benefits in a near-real time fashion.

PPR builds on existing encapsulations at the edge nodes to add path identity to the packet. This reduces the per packet overhead required for path steering and therefore has a smaller impact on both packet MTU, data plane processing and overall goodput for small payload packets, while extending path steering benefits to any existing data plane.

A number of extensions that allow expansion of use beyond simple point-to-point-paths are also described in this memo.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of [BCP 78](#) and [BCP 79](#).

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <https://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any

time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on April 28, 2022.

Copyright Notice

Copyright (c) 2021 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to [BCP 78](#) and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1.	Introduction	3
1.1.	Relation to Segment Routing	4
1.2.	Requirements Language	4
1.3.	Acronyms	4
2.	Applicability and Key use cases	5
2.1.	XHaul Transport	6
2.2.	PPR as VPN+ Underlay and Network Slicing	7
2.3.	PPR as FRR Solution	7
2.4.	Extensible alternative to Flex Algo	8
2.5.	PPR for energy-optimized networks	8
3.	Preferred Path Routing (PPR)	9
3.1.	PPR Data Plane aspects	11
3.1.1.	PPR Native IP Data Planes	11
3.1.2.	SR-MPLS with PPR	13
3.1.3.	SRv6, Network Programming and PPR	13
3.2.	PPR Control Plane aspects	14
3.2.1.	PPR-ID and Data Plane Extensibility	14
3.2.2.	PPR Path Description Elements (PDEs)	14
3.2.3.	ECMP Considerations	15
3.2.4.	PPR Services along the Path	15
3.2.5.	PPR Graphs	15
3.2.6.	PPR Multi-Domain Scenarios	17
3.3.	PPR Management Plane Aspects	18
3.3.1.	IGP Metric Independent Paths/Graphs	18
3.3.2.	Granular OAM	19
4.	Preferred Path Loop Free Alternatives (pLFA)	20

5.	Traffic Engineering Attributes	21
6.	IANA Considerations	22
7.	Security Considerations	22
8.	References	22
8.1.	Normative References	22
8.2.	Informative References	22
	Authors' Addresses	24

[1.](#) Introduction

With the deployments of more advanced services, such as 5G and beyond, the need for Traffic Engineering (TE) with deterministic services become more important. Especially in many edge networks where stringent requirements must be met in terms of latency, throughput, packet loss and packet error rate. Traffic steering provides a base to build some of these capabilities to serve various cellular, edge and vertical industries. Additionally, diverse data planes are used in various deployments and parts of the network, including Ethernet, MPLS, and native IP (IPv4/IPv6) needs some of these capabilities.

This document provides a framework for Preferred Path Routing (PPR). PPR is a method of adding explicit paths to a network using link-state routing protocols. Such a path, which may be a strict or loose and can be any loop-free path between two points in the network. A node makes an on-path check to determine if it is on the path, and, if so, adds a FIB entry with NextHop (NH) (computed from the SPF tree) set to the next element in the path description.

The Preferred Path Route Identifier (PPR-ID) in the packet is used to map the packet to the PPR path, and hence to identify resources and the NH. In other words, PPR-ID is the path identity to the packet and routing and forwarding happens based on this identifier while providing various services to all the flows mapped to the path.

As described, PPR is forwarding plane agnostic, and may be used with any packet technology in which the packet carries an identifier that is unique within the PPR domain. PPR may hence be used to add explicit path and resource mapping functionality with inherent TE properties in IPv4, IPv6, MPLS, Ethernet or similar networks. It also has a smaller impact on both packet MTU and data plane processing. PPR uses an IGP control plane based approach for dynamic path steering.

Applications and key use case scenarios for PPR are described further in [Section 2](#).

PPR itself is described in further detail in [Section 3](#).

[Section 3.1](#), [Section 3.2](#), [Section 3.3](#) describe various data, control and management plane functionalities of PPR respectively. A number of extensions that allow expansion of use beyond simple point-to-point- paths, TE aware loop-free-alternatives and path related services to the flows are also described in this memo.

[1.1.](#) Relation to Segment Routing

Segment Routing (SR) [[RFC8402](#)] enables packet steering by including set of Segment Identifiers (SIDs) in the packet that the packet must traverse or be processed by. In an MPLS network this is done by mapping the SIDs to MPLS labels and then pushing the required labels on the packet [[RFC8660](#)]. In SRv6, [[RFC8754](#)] defines a segment routing extension header (SRH) to be carried in the packet which contains a list of the segments. The usefulness of PPR with SR and inter- working scenarios are described in [Section 3.1.2](#) and [Section 3.1.3](#).

SR also defines Binding SIDs (BSIDs) [[RFC8402](#)] which are SIDs pre-positioned in the network to either allow the number of SIDs in the packet to be reduced, or provide a method of translating from an edge imposed SID to a SID that the network prefers. One use of BSIDs is to define a path by associating an out-bound SID on every node along the path in which case the packet can be steered by swapping the incoming active SID on the packet with a BSID. For both SR-MPLS and SRv6, PPR can reduce the number of touch points needed with BSIDs by dynamically signaling the path and associating the path with an abstract data plane identifier.

With PPR as a data packet carries a PPR-ID [Section 3.1](#) instead of individual SIDs, it avoids exposing the path; thus it avoids revealing topology, traffic flow and service usage, if a packet is snooped. This is described as "Topology Disclosure" security consideration in [[RFC8754](#)].

[1.2.](#) Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119](#) [[RFC2119](#)], [RFC8174](#) [[RFC8174](#)] when, and only when they appear in all capitals, as shown here.

[1.3.](#) Acronyms

Term	Definition
GTP	GPRS Tunneling Protocol (3GPP)
IS-IS LSP	IS-IS Link State PDU
IPFRR	IP FastReRoute
MPLS	Multi-Protocol Label Switching
MTU	Maximum Transferable Unit
NH	NextHop
PDE	Path Description Element of the Preferred path
PE	Provider Edge
PPR	Preferred Path Routing/Route
PPR-ID	Preferred Path Route Identifier, a data plane identifier
(R)AN	5G Radio Access Network (3GPP REL15)
SID	Segment Identifier
SPF	Shortest Path First
SR-MPLS	Segment Routing with MPLS data plane
SRH	Segment Routing Header - IPv6 routing Extension header
SRv6	Segment Routing with IPv6 data plane with SRH
TE	Traffic Engineering
UPF	User Plane Function (3GPP REL15)

2. Applicability and Key use cases

2.1. XHaul Transport

Cellular networks predominantly use both IP and MPLS data planes in the transport part of the network. For the cellular transport to evolve for 5G, certain underlay network requirements need to be met (for slices other than enhanced Mobile Broadband (eMBB)). PPR is a mechanism to achieve this, as it provides dynamic path based routing and traffic steering for any underlying data plane (IPv4/IPv6/MPLS) used, without any additional control plane protocol in the network. PPR acts as an underlay mechanism in cellular XHaul (N3/N9 interfaces below) and hence can work with any overlay mechanism including GPRS Tunneling Protocol (GTP).

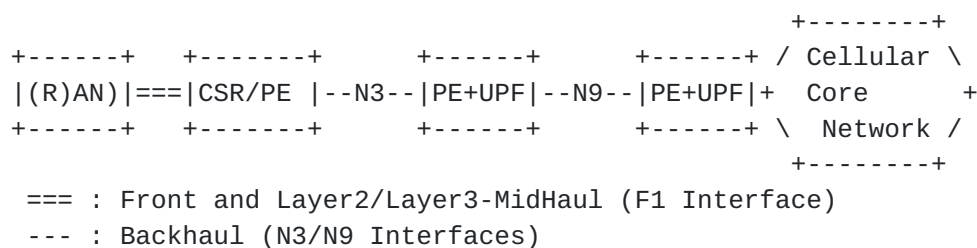


Figure 1: Cellular Transport Network

Figure 1 depicts a high level view of cellular XHaul network. Fronthaul is generally a point-to-point link, midhaul uses Layer-2/IP and backhaul is an IP/MPLS network. For the end-to-end slicing in these deployments, both midhaul and backhaul need to have traffic engineering as well as underlay QoS capabilities.

In many cellular deployments connectivity for various 5G nodes on F1, N3 and N9 interfaces, topologies for example, range from subtended rings to Leaf- Spine Fabric (LS-Fabric) in edge deployments. While there is no limitation w.r.t topologies where PPR is applicable, for some of these deployments PPR is more suitable for providing Traffic Engineering for high volume traffic with no path overhead in the underlying data plane. PPR augments the SR-MPLS deployments with low data plane overhead and high reliability with TE aware fast reroute (PLFA) as described in [Section 3.2.2](#). In the overlay or virtual router environment, PPR provides lightweight service chaining with non-topological Path Description Element (PDEs) [Section 3.2.2](#) along the preferred path. PPR helps to achieve OAM capabilities at the path granularity without any additional per packet information.

Some edge deployment underlays are predominantly IP (IPv4/IPv6) based. If IGP based underlay control plane is in use, PPR can provide the required flexibility for creating TE paths, where native IP data planes (IPv4/IPv6) are used. PPR can help operators to

mitigate the congestion in the underlay and path related services for critical servers in the edge networks dynamically.

2.2. PPR as VPN+ Underlay and Network Slicing

There is a need to support the requirements of new applications, particularly applications that are associated with 5G services. An approach to supporting these needs is described in [\[I-D.ietf-teas-enhanced-vpn\]](#). This approach utilizes existing VPN and TE technologies and adds features that specific services require over and above traditional VPNs. The document describes a framework for using existing, modified and potential new networking technologies as components to provide an Enhanced Virtual Private Network (VPN+) service.

Typically, VPN+ will be used to form the underpinning of network slicing, but could also be of use in its own right. It is not envisaged that large numbers of VPN+ instances will be deployed in a network and, in particular, it is not intended that all VPNs supported by a network will use VPN+ techniques.

Such networks potentially need large numbers of paths each with individually allocated resources at each link and node. A segment routing approach has the potential to require large numbers of SIDs in each packet; and the paths become strict source routed through end-to-end set of resources needed to create the VPN+ paths. By using PPR, the number of segments needed in packets is reduced, and the management overhead of installing the large numbers of BSIDs is reduced.

2.3. PPR as FRR Solution

PPR may be used in a network as a method of providing IP Fast-ReRoute (IPFRR). This is independent of whether PPR is used in the network for other traffic steering purposes. It can be used to create optimal paths or paths congruent with the post convergence path from the point of local repair (PLR) as is proposed in TI-LFA [\[I-D.ietf-rtgwg-segment-routing-ti-lfa\]](#). Unlike TI-LFA PPR may be used in IPv4 networks. This is discussed further in [Section 4](#). The approach has the further intrinsic advantage that no matter how complex the repair path, only a single header (or MPLS label) needs to be pushed onto the packet which may assist routers that find it difficult to push large headers.

2.4. Extensible alternative to Flex Algo

Flex-Algorithm [[I-D.ietf-lsr-flex-algo](#)] is a method that is sometimes used to create paths between Segment Routing (SR) nodes when it is required that packets traverse a path other than the shortest path that the SPF of the underlying IGP would naturally install. There is a limit of 128 algorithms that can be installed in a network. Flex-Algorithm is a cost based approach to creating a path which means that a path or pathlet is indirectly created by manipulating the metrics of the links. These metrics affect all the paths within the scope of the Flex-Algorithm number (instance). The traffic steering properties of Flex-Algorithm required for SR can be achieved directly with PPR with several advantages:

- o The scope of a PPR path is strictly limited to the sub-path between the SR nodes.
- o The path can be directly specifies rather than implicitly through metrics
- o Resources (such as specialist queues etc.) may be directly mapped to the PPR path and hence to the SR sub-path.

2.5. PPR for energy-optimized networks

Improving energy efficiency and reducing power consumption are becoming of increasing importance for many industries, which includes network operators as well as users and providers of networking services. Network providers can contribute to addressing those challenges by making their networks more energy-efficient. This includes the support of power saving schemes that guide traffic along paths deemed particularly "energy efficient".

A significant opportunity to reduce power consumption concerns powering down (or putting into power saving mode) equipment (including line card, ports, etc.) that may not be currently needed. At the same time, the incremental power consumption for additional traffic on ports and equipment already under power is for all practical purposes negligible. Optimizing energy efficiency thus involves directing traffic in such a way that it allows for isolation of equipment that may at the moment not be needed so that it could be powered down or brought into power-saving mode.

This implies that power efficiency of networks can be greatly affected by the paths along which traffic is directed at any particular point in time. Proper engineering of those paths and the ability to configure them effectively thus becomes an important tool to optimize power usage of networks and make them more energy

efficient. PPR provides a mechanism that enables this, allowing to engineer dynamic paths that optimize the network from an energy savings perspective as one important set of criteria for any underlying data plane in the network.

3. Preferred Path Routing (PPR)

PPR allows the direction of traffic along an engineered path through the network by replacing the SID label stack or the SID list with a single PPR-ID. The PPR-ID may either be a single label (MPLS) or a native destination prefix (IPv4/IPv6). This enables the use of a single data plane identifier to describe an entire path.

A PPR path could be an (Segmented Routed) SR path, a traffic engineered path computed based on some constraints, an explicitly provisioned Fast Re-Route (FRR) path or a service chained path. A PPR path can be signaled by any node, computed by a central controller, or manually configured by an operator. PPR extends the source routing and path steering capabilities to native IP (IPv4 and IPv6) data planes without hardware upgrades; see [Section 3.1](#).

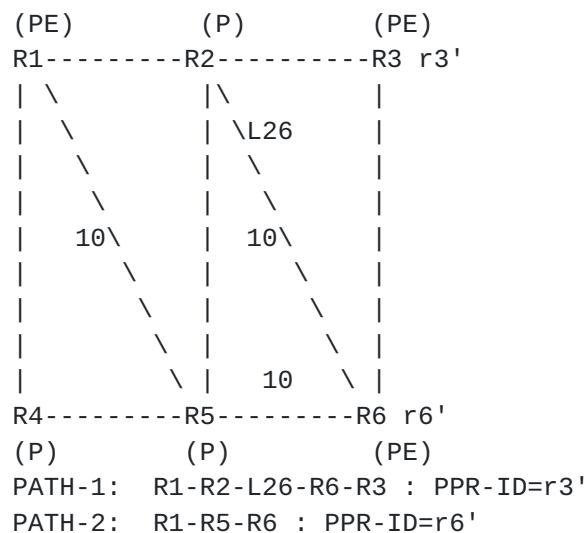


Figure 2: IGP Network

In the Figure 2, consider node R1 as an ingress node, or a head-end node, and the node R3 may be an egress node or another head-end node. The numbers shown on links between nodes indicate the bi-directional IGP metric as provisioned (no number indicates a metric of 1). R1 may be configured to receive TE source routed path information from a central entity (PCE [[RFC5440](#)], Netconf [[RFC6241](#)] or a Controller) that comprises PPR information which relates to sources that are attached to R1. It is also possible to have a PPR provisioned

locally by the operator for non-TE needs (e.g., FRR or for chaining certain services).

The PPR is encoded as an ordered list of path elements from source to a destination node in the network and is represented with a PPR-ID to represent the path. The path can represent both topological and non-topological elements (for example, links, nodes, queues, priority and processing actions) and specifies the actual path towards the egress node.

- o The shortest path towards R3 from R1 is through the following sequence of nodes: R1-R2-R3 based on the provisioned IGP metrics.
- o The central entity in this example, can define a PPRs from R1 to R3 and R1 to R6 that deviate from the shortest path based on other network characteristic requirements as requested by an application or service. For example, the network characteristics or performance requirements may include bandwidth, jitter, latency, throughput, error rate, etc.
- o In a VPN setup, nodes R1, R3 and R6 are PE nodes and other nodes are P nodes. User traffic entering at the ingress PE nodes gets encapsulated (e.g., MPLS, GRE, GTP, IP-IN-IP, GUE) and will be delivered to the egress PE.

Consider two paths in the above network:

- o PATH-1: A first PPR may be identified by PPR-ID = r3' with the path description R1-R2-L26-R6-R3 for a Prefix advertised by R3. This is an example of a strict path with a combination of links and nodes.
- o PATH-2: A second PPR may be identified by PPR-ID = r6' with the path description R1-R5-R6. This is an example of a loose path. Though this example shows PPRs with node identifiers it is possible to have a PPR with a combination of Non-Topological elements along the path.

The first topological element relative to the beginning of PPR Path descriptor contains the information about the first node in the path that the packet must pass through (e.g. equivalent to the top label in SR-MPLS and the first SID in an SRv6 SRH). The last topological sub-object or PDE contains information about the last node (e.g. in SR-MPLS it is equivalent to the bottom SR label).

Each IGP node receiving a complete path description, determines whether the node is on the advertised PPR path. This is called the PPR on-path check. It then determines whether it is included more

than once on that path. This PPR validation prevents the formation of a routing loop. If the path is looped, no further processing of the PPR is undertaken. (Note that even if it is invalid, the PPR descriptor must still be flooded to preserve the consistency of the underlying routing protocol). If the validation succeeds, the receiving IGP node installs a Forwarding Information dataBase (FIB) entry for the PPR-ID with the NextHop (NH) required to take the packet to the next topological path element in the path description. Processing of PPRs may be done at the end of the IGP SPF computation.

Consider PPR path PATH-1 in Figure 2. When node R5 receives the PPR (PATH-1) information it does not install a FIB entry for PATH-1 because this PPR does not include node R5 in the path description/ordered path list.

However, node R5 determines that the second PPR (PATH-2), does include the node R5 in its path description (the on-path check passes). Therefore, node R5 updates its FIB to include an entry for the destination address that R6 indicates (PPR-ID) along with path description. This allows the forwarding of data packets with the PPR-ID (r6') to the next element along the path, and hence towards node R6.

To summarize the control plane processing, the receiving IGP node determines if it is on the path by checking the node's topological elements in the path list. If it is, it adds/adjusts the PPR-ID's shortest path NH towards the next topological path element in the PPR's path list. This process continues at every IGP node as specified in the path description TLV.

3.1. PPR Data Plane aspects

Data plane type for PPR-ID is selected by the entity (e.g., a controller, locally provisioned by operator), which selects a particular PPR in the network.

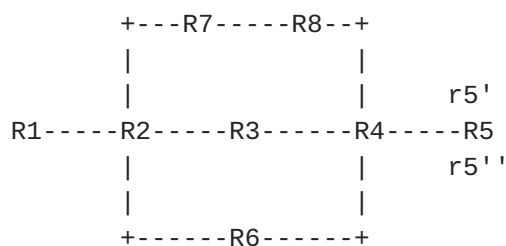
3.1.1. PPR Native IP Data Planes

In an IPv4 network, source routing and packet steering with PPR can be done by selecting the IPv4 data plane type (PPR-IPv4), in PPR Path description with a corresponding IPv4 address/prefix as PPR-ID while signaling the path description in the control plane [Section 3.2](#). Forwarding is done by setting the destination IP address of the packet as PPR-ID at the ingress node of the network. In this case this is an IPv4 address in the tunneled/encapsulated user packet. There is no data plane change or upgrade needed to support this functionality.

Similarly, for an IPv6 network source routing and packet steering can be done in IPv6 data plane type (PPR-IPv6), along the path as described in PPR Path description with a corresponding IPv6 address/prefix as PPR-ID in the control plane [Section 3.2](#). Whatever specified above for IPv4 applies here too, except that destination IP address of the encapsulated data packet at the edge node is an IPv6 Address (PPR-ID). This doesn't require any IPv6 extension headers (EH).

For a loose path in an IPv4 or IPv6 network (Native IPv4 or IPv6 data planes respectively) the packet has to be encapsulated using the capabilities (either dynamically signaled through [\[I-D.ietf-isis-encapsulation-cap\]](#) or statically provisioned on the nodes) of the next loose PDE in the path description.

Consider the network fragment shown in Figure 3 which further illustrates loose routing, and consider PATH-3. Node R2 can reach R5 ECMP through R2->R3->R4, and R2->R6->R4, both at cost 2. The path R2->R7->R8->R4 is longer (cost 3) and is not a path that R2 would choose to use it to reach R4. Node R2 (start of the loose segment) is programmed to encapsulate a data packet towards the next loose topological PPR-PDE in the path, which is R4. The NH computed at R1 (for PPR-ID r5') would be the shortest path towards R5 i.e., the interfaces towards R2. R2 has an ECMP towards R3 and R6 to reach R4 (next PDE in the loose segment), as packet would be encapsulated at R2 for R4 as the destination. R7 and R8 are not involved in this PPR path and so do not need a FIB entry for PPR-ID r5' (the on-path check for PATH-3 fails at these nodes).



All costs are 1

PATH-3: R1-R2-R4-R5 : PPR-ID=r5'

PATH-4: R1-R2-R3-R4-R5 : PPR-ID=r5''

Figure 3: Network with Loose Path

In a strict path, for example, PATH-4 in Figure 3, PPR-ID is programmed on the data plane at each node of the path, with NH set to the shortest path towards the next topological PPR-PDE. In this case, no further encapsulation of the data packet is required.

3.1.2. SR-MPLS with PPR

PPR is fully backward compatible with the SR data plane. As control plane PDEs can be extensible and particular data plane identifiers can be expressed to describe the path, in SR case PDEs can contain the SR SIDs.

In SR-MPLS, a data packet contains the stack of labels (path steering instructions) which guides the packet traversal in the network. For SR-MPLS data plane, the complete set of label stack is represented with a unique SR SID/Label, PPR-ID, to represent the path. The PPR-ID gets programmed on the data plane of each node, with the appropriate NH computed as specified in [Section 3](#). PPR-ID here is a label/index from the SRGB (like another node SID or global ADJ-SID). PPR path description in the control plane is a set of ordered SIDs represented with PPR-PDEs. Non-Topological segments described along with the topological PDEs can also be programmed in the forwarding plane to enable specific function/service, when the data packet hits with corresponding PPR-ID.

For SR-MPLS data plane, either 1 label or 2 labels need to be provisioned on individual nodes on the path description. In the example network Figure 2, for PATH-2 (a loose path), during control plane processing, node R1 programs the bottom label as PPR-ID and the top label as the next topological PPR-PDE in the path, which is a node SID of R5. In the control plane, the NH computed at R1 would be the shortest path towards R5 i.e., the interfaces towards R2 and R4 (ECMP). For strict paths, a single label (PPR-ID) is programmed on the data plane along the path, with NH set to the shortest path towards the next topological PPR-PDE in the path description.

3.1.3. SRv6, Network Programming and PPR

One of the key benefits PPR offers for SRv6 data plane is an optimized data plane as individual path steering SIDs in the data packet is replaced with a path identifier (PPR-ID). Thus potentially avoids MTU, hardware incompatibilities and processing overhead. Few PPR and SRv6 inter working scenarios are listed below.

In a simple encapsulation mode without SRH [[RFC8754](#)], an SRv6 SID can be used as PPR-ID. With this approach path steering can be brought in with PPR and some of the network functions as defined in [[RFC8986](#)] can be realized at the egress node as PPR-ID in this case is a SRv6 SID.

In SRv6 with SRH, one-way PPR-ID can be used, by setting it as the destination IPv6 address and SL field in SRH is set to 0; here, SRH can contain any other TLVs and non-topological SIDs as needed.

Another inter working case can be a multi area IGP deployment. In this case multiple PPR-IDs corresponding to each IGP area can be encoded as SIDs in SRH for an end-to-end path steering with minimal SIDs in SRH.

[3.2.](#) PPR Control Plane aspects

[3.2.1.](#) PPR-ID and Data Plane Extensibility

The data plane identifier, PPR-ID, describes a path through the network. A data plane type and corresponding PPR-ID can be specified with the advertised path description in the IGP. The PPR-ID type allows data plane extensibility for PPR, though it is currently defined for IPv4, IPv6, SR-MPLS and SRv6 data planes.

For native IP data planes, this is mapped to either IPv4 or IPv6 address/prefix. For SR-MPLS, PPR-ID is mapped to an MPLS Label/SID and for SRv6, this is mapped to an IPv6-SID. This is further detailed in [Section 3.1](#) and [Section 3.1.3](#).

[3.2.2.](#) PPR Path Description Elements (PDEs)

The path identified by the PPR-ID is described as a set of Path Description Elements (PDEs), each of which represents a segment of the path. Each node determines its location in the path as described, and forwards to the next segment/hop or label of the path description (see the Forwarding Procedure Example later in this document).

These PPR-PDEs like SR SIDs, can represent topological elements like links/nodes, backup nodes, as well as non- topological elements such as a service, function, or context on a node with additional control information as needed.

A preferred path can be described as a Strict-PPR or a Loose-PPR. In a Strict-PPR all nodes/links on the path are described with SR-SIDs for SR data planes or IPv4/IPv6 addresses for native IP data planes. In a Loose-PPR only some of the nodes/links from source to destination are described. More specifics and restrictions around Strict/Loose PPRs are described in respective data planes in [Section 3.1](#) and [Section 3.1.3](#). Each PDE is described as either an MPLS label towards the NH in MPLS enabled networks, or as an IP NH, in the case of either 'plain'/'native' IP or SRv6 enabled networks. A PPR path is related to a set of PDEs using the TLVs specified in the respective IGPs.

3.2.3. ECMP Considerations

PPR inherently supports Equal Cost Multi Path (ECMP) for both strict and loose paths. If a path is described using nodes, it would have ECMP NHs established for PPR-ID along the path. In the network shown in Figure 2, for PATH-2, node R1 would establish ECMP NHs computed by the IGP, towards R5 for the PPR-ID r6'. However, one can avoid ECMP on any segment of the path by pinning the path using link identifier to the next segment as specified for PATH-1 in Figure 2.

3.2.4. PPR Services along the Path

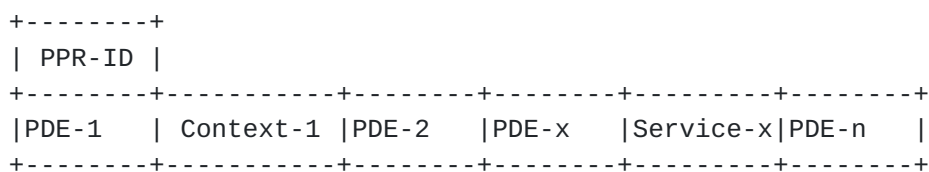


Figure 4: Services along the Preferred Path

As shown in Figure 4, some of the services specific to a preferred path, can be encoded as non-topological PDEs and can be part of the path description. These services are applied at the respective nodes along the path. In Figure 4, PDE-1, PDE-2, PDE-x, PDE-n are topological PDEs of a data plane. For SR-MPLS/SRv6 data planes these are simply SIDs and for native IP data planes corresponding non-topological addresses. When the data packet with a PPR-ID is delivered to node-1, the packet is delivered to Context-1. Similarly on node-x, Service-x is applied. These services/functions need to be pre-provisioned on the particular nodes and optionally can be advertised in IGPs.

The above gives the basic and light weight service chaining capability with PPR without incurring any additional overhead on the data packet. However, this is limited to fixed services/functions for a path and all data packets using the path will be applied with these services. Flow level exclusions using the same path or differentiated services that need to be applied within a flow cannot be supported with this mechanism and one has to resort to data plane mechanisms as defined in NSH/SFC [[RFC8300](#)].

3.2.5. PPR Graphs

In a network of N nodes a total $O(N^2)$ unidirectional paths are necessary to establish any-to-any connectivity, and multiple (k) such path sets may be desirable if multiple path policies are to be supported (lowest latency, highest throughput etc.).

Each PPR Tree uses one label/SID and defines paths from any set of nodes to one destination, this reduces the number of entries needed. For example, it reduces the number of forwarding identifiers needed in SR-MPLS data plane [Section 3.1.2](#) with PPR, which are derived from the SRGB at the egress node. These paths are form a tree rooted at the destination. In other word, PPR Tree identifiers are destination identifiers and PPR Trees are path engineered destination routes

(like IP routes) and its scaling simplifies to linear in N i.e., $O(k*N)$.

In a completely different usage paradigm, a PPR Graph can also have multiple forwarding identifiers (PPR-IDs). Based on the algorithm specified for the Graph, path computation can be done in a distributed fashion in the network to establish the forwarding over the graph. Various types of PPR Graphs, rules for construction and their usage details will be described in future revisions.

3.2.6. PPR Multi-Domain Scenarios

PPR can be extended to multi-domain, including multi-area scenarios as shown in Figure 6.

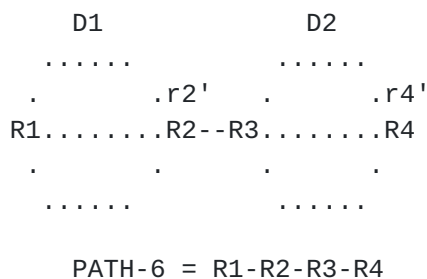


Figure 6: Multi-Domain Network with PPR

Operation of PPR within the domain is as described in the preceding sections of this document. The key difference in operation in multi-domain concerns the value of the PPR-ID in the packet. There are three approaches that can be taken:

1. The PPR-ID is constant along the end-end-path. This requires coordination of the PPR-ID in each domain. This has the convenience of a uniform identity for the path. However, whilst an IPv6 network has a large PPR identity space, this is not the case for MPLS and is less the case for IPv4. The approach also has the disadvantage that the entirety of the domains involved need to be configured and provisioned with the common value. In the network shown in Figure 6 The PPR-ID for PATH-6 is $r4'$.
2. The PPR-ID for each individual domain is the value that best suits that domain, and the PPR-ID is swapped at the boundary of the domains. This allows a PPR-ID that best suits each domain. This is similar to the approach taken with multi-segment pseudowire [RFC5659]. This approach better suits the needs of network layers with limited identity resources. It also enables the better coordination of PPR-IDs. In this approach the PPR-ID for PATH-6 would be $r2'$ in domain D1 and $r4'$ in domain D2. These

two PPR-IDs would be distributed in their own domains and the only inter-domain co-ordination required would be between R2 and R3.

3. A variant of (2) is that the PPR-IDs are domain specific, but a segment routing approach is taken in which they encoded at ingress (R1), and are popped at the inter-domain boarder. This requires that the domain ingress and egress routers support segment routing data-plane capability.

Although the example shown in Figure 6 shows the case of two domains, nothing limits the design to just two IGP areas. This further explained below.

In controller based deployments, each IGP area can have separate north bound and south bound communication end points with PCE/SDN controller, in their respective domain. It is expected that PPR paths for each IGP level are computed and provisioned at the ingress nodes of the corresponding area's area boarder router. Separate path advertisement in the respective IGP area should happen with the same PPR-ID. With this, only PPR-ID needs to be leaked to the other area, as long as a path is available in the destination area for that PPR-ID. If the destination area is not provisioned with path information, area boarder shall not leak the PPR-ID to the destination area.

3.3. PPR Management Plane Aspects

3.3.1. IGP Metric Independent Paths/Graphs

PPR allows a considerable simplification in the design and management of networks. In a best effort network the setting of the IGP metrics is a complex problem with competing constraints. A set of metrics the is optimal for traffic distribution under normal operation may not be optimal under conditions of failure of one or more of the network components. Nor is that choice of metrics necessarily best for operation under all IPFRR conditions. When SR is introduced to the network a further constraint on metrics is the need to limit the size of the SID stack/list. These problems further increase with the introduction of demanding technologies such as network slicing and deterministic networking.

Some mitigation occurs with the use of FlexAlgo [[I-D.ietf-lsr-flex-algo](#)] but fundamentally this is still an approach that is critically dependent on the per-flex-algo provisioning of different metrics on participating nodes, that operate in both the normal and the failure case.

PPR allows the network to simply introduce metric independent paths on a strategic or tactical basis. Being metric independent each PPR path operates ships-in-the-night with respect to all other paths. This means that the network management system can address network tuning on a case by case basis only needing to worry about the traffic matrix along the path rather than needing to deconvolve the impact of tuning a metric on the whole traffic matrix. In other words, PPR is a direct method of tuning the traffic rather than an the indirect method that metric tuning provides.

An example that makes this clear is the maximally redundant tree (MRT) approach to IPFRR. MRT requires the tuning of metrics to tune the paths, and a common algorithm for all nodes in the network. An equivalent solution can be introduced to the network by the insertion of a pair of PPR graphs by the network management system. Furthermore the topology of these graphs are independent of all other graphs, allowing the tuning and migration of the repair paths in the network management system.

Thus PPR allows the operator to focus on the desired traffic path of specific groups of packets independent of the desired path of the packets in all other paths.

3.3.2. Granular OAM

For some of the deployments as described in [Section 2](#), the ability to collect certain statistics about PPR path usage, including how much traffic a PPR path carries and at what times from any node in the network is a critical requirement. Such statistics can be useful to account for the degree of usage of a path and provide additional operational insights, including usage patterns and trending information.

Traffic for certain PPRs may have more stringent requirement w.r.t accounting for critical SLAs (e.g. 5G non-eMBB slice) and should account for any link/node failures along the path. Optional per path attributes like "Packet Traffic Accounting" and "Traffic Statistics" instructs all the respective nodes along the path to provision the hardware and to account for the respective traffic statistics. Traffic accounting should be applied based on the PPR-ID. This capability allows a more granular and dynamic measurement of traffic statistics for only certain PPRs as needed.

As routing happens on the abstracted path identifier in the packet, no additional per packet instruction is needed for achieving the above functionality regardless of the data plane used in the network [Section 3.1](#).

4. Preferred Path Loop Free Alternatives (pLFA)

PPR can be used as a method of providing IPFRR. Preferred Path Loop-Free Alternate (pLFA) allows the construction of arbitrary engineered backup paths pLFA and inherits the low packet overhead of PPR requiring a simple encapsulation and a single path identifier for any path of any complexity.

pLFA provides a superset of RSVP-TE repairs (complete with traffic engineering capability) and Topology Independent Loop-Free Alternates (TI-LFA) [[I-D.ietf-rtgwg-segment-routing-ti-lfa](#)]. However, unlike the TI-LFA approaches PPR is applicable to a more complete set of data planes (for example MPLS, both IPv4 and IPv6 and Ethernet) where it can provide a rich set of IPFRR capabilities ranging from simple best-effort repair calculated at the point of local repair (PLR) to full traffic engineered paths. For any repair path pLFA requires one encapsulation and one PPR-ID, regardless of the complexity and constraints of the path.

For a basic understanding of pLFA consider the case of a link repair shown in this example as shown in Figure 7, we assume that we have a path A-B-C-D that the packet must traverse. This may be a normal best effort path or a traffic engineered path.

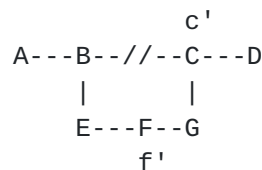


Figure 7

PPR is used to inject the repair path B->E->F->G->C into the network with a PPR-ID of c'. B is monitoring the health of link B->C, for example looking for loss-of-light, or using Bidirectional Forwarding Detection (BFD) [[RFC5880](#)]. When B detects a failure it encapsulates the packet to C by adding to the packet an encapsulation with a destination address set as the PPR-ID for c' and then sending the packet to E. At C the packet is decapsulated and sent to D. The path B->E->F->G->C may be a traffic engineered path or it may be a best effort path. This may of course be the post convergence path from B to C, as is used by TI-LFA However B may have at its disposal multiple paths to C with different properties for different traffic classes. In this case each path to be used would require its own PPR-ID (c', c'' etc.). Because pLFA only requires a single path identifier regardless of the complexity of the path is not necessary constrain the path to be a small number of loose source routed paths to protect against MTU or maximum SID count considerations.

pLFA supports the usual IPFRR features such as early release into Q-space, node repair, and shared risk link group support, LANs, ECMP and multi-homed prefixes. However the ability to apply repair graphs [Section 3.2.5](#) is unique to pLFA. The use of graphs in IPFRR repair simplifies the construction of traffic engineered repair paths, and allows for the construction of arbitrary maximally redundant tree repair paths.

Of importance in any IPFRR strategy in a loosely routed network, including normal connectionless routing is the ability to support loop-free convergence. This problem is described in [\[RFC5715\]](#). [\[I-D.ietf-rtgwg-segment-routing-ti-lfa\]](#) has proposed a mitigation technique for failures (noted above) and pLFA is able to support this. However a network supporting high reliability traffic may find mitigation insufficient. Also disruption can take place on network component inclusion (or repair/recovery) and TI-LFA is silent on this. A network using pLFA is compatible with all of the know loop-free convergence and loop mitigation approaches described in [\[RFC5715\]](#).

5. Traffic Engineering Attributes

In addition to determining the nodes to traverse, there may be other aspects that need to be set up for a path. Most notably, this concerns the allocation and reservation of resources along the path to help ensure the service levels, i.e. the Quality of Service that is delivered across the path, will be acceptable for the traffic routed across the path (critical in some deployments as listed in [Section 2](#)).

While SR allows packet steering on a specified path (for MPLS and IPv6 with SRH), it does not have any notion of QoS or resources reserved along the path. The determination of which resources to allocate and reserve on nodes across the path, like the determination of the path itself, can in many cases be made by a controller. Accordingly, PPR includes extensions that allow to manage those reservations, in addition to the path itself.

Key aspect of the solution concerns with specifying the resources to be reserved along the preferred path, through path attributes TLVs. Reservations are expressed in terms of required resources (bandwidth), traffic characteristics (burst size), and service level parameters (expected maximum latency at each hop) based on the capabilities of each node and link along the path. The second part of the solution is providing mechanism to indicate the status of the reservations requested i.e. if these have been honored by individual node/links in the path. This can be done by defining a new TLV/Sub-TLV in respective IGPs. Another aspect is additional node level TLVs

and extensions to IS-IS-TE [[RFC7810](#)] and OSPF-TE [[RFC7471](#)] to provide accounting/usage statistics that have to be maintained at each node per preferred path.

6. IANA Considerations

This document does not request any allocations from IANA.

7. Security Considerations

Advertisement of the additional information defined in this document introduces no new security concerns in IGP protocols. However, for extensions related to SR-MPLS and SRH data planes, those particular data plane security considerations does apply here.

8. References

8.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), DOI 10.17487/RFC2119, March 1997, <<https://www.rfc-editor.org/info/rfc2119>>.
- [RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in [RFC 2119](#) Key Words", [BCP 14](#), [RFC 8174](#), DOI 10.17487/RFC8174, May 2017, <<https://www.rfc-editor.org/info/rfc8174>>.

8.2. Informative References

- [I-D.ietf-isis-encapsulation-cap]
Xu, X., Decraene, B., Raszuk, R., Chunduri, U., Contreras, L. M., and L. Jalil, "Advertising Tunnelling Capability in IS-IS", [draft-ietf-isis-encapsulation-cap-01](#) (work in progress), April 2017.
- [I-D.ietf-lsr-flex-algo]
Psenak, P., Hegde, S., Filsfils, C., Talaulikar, K., and A. Gulko, "IGP Flexible Algorithm", [draft-ietf-lsr-flex-algo-18](#) (work in progress), October 2021.
- [I-D.ietf-rtgwg-segment-routing-ti-lfa]
Litkowski, S., Bashandy, A., Filsfils, C., Francois, P., Decraene, B., and D. Voyer, "Topology Independent Fast Reroute using Segment Routing", [draft-ietf-rtgwg-segment-routing-ti-lfa-07](#) (work in progress), June 2021.

- [I-D.ietf-teas-enhanced-vpn]
Dong, J., Bryant, S., Li, Z., Miyasaka, T., and Y. Lee, "A Framework for Enhanced Virtual Private Network (VPN+) Services", [draft-ietf-teas-enhanced-vpn-09](#) (work in progress), October 2021.
- [RFC5440] Vasseur, JP., Ed. and JL. Le Roux, Ed., "Path Computation Element (PCE) Communication Protocol (PCEP)", [RFC 5440](#), DOI 10.17487/RFC5440, March 2009, <<https://www.rfc-editor.org/info/rfc5440>>.
- [RFC5659] Bocci, M. and S. Bryant, "An Architecture for Multi-Segment Pseudowire Emulation Edge-to-Edge", [RFC 5659](#), DOI 10.17487/RFC5659, October 2009, <<https://www.rfc-editor.org/info/rfc5659>>.
- [RFC5715] Shand, M. and S. Bryant, "A Framework for Loop-Free Convergence", [RFC 5715](#), DOI 10.17487/RFC5715, January 2010, <<https://www.rfc-editor.org/info/rfc5715>>.
- [RFC5880] Katz, D. and D. Ward, "Bidirectional Forwarding Detection (BFD)", [RFC 5880](#), DOI 10.17487/RFC5880, June 2010, <<https://www.rfc-editor.org/info/rfc5880>>.
- [RFC6241] Enns, R., Ed., Bjorklund, M., Ed., Schoenwaelder, J., Ed., and A. Bierman, Ed., "Network Configuration Protocol (NETCONF)", [RFC 6241](#), DOI 10.17487/RFC6241, June 2011, <<https://www.rfc-editor.org/info/rfc6241>>.
- [RFC7471] Giacalone, S., Ward, D., Drake, J., Atlas, A., and S. Previdi, "OSPF Traffic Engineering (TE) Metric Extensions", [RFC 7471](#), DOI 10.17487/RFC7471, March 2015, <<https://www.rfc-editor.org/info/rfc7471>>.
- [RFC7810] Previdi, S., Ed., Giacalone, S., Ward, D., Drake, J., and Q. Wu, "IS-IS Traffic Engineering (TE) Metric Extensions", [RFC 7810](#), DOI 10.17487/RFC7810, May 2016, <<https://www.rfc-editor.org/info/rfc7810>>.
- [RFC8300] Quinn, P., Ed., Elzur, U., Ed., and C. Pignataro, Ed., "Network Service Header (NSH)", [RFC 8300](#), DOI 10.17487/RFC8300, January 2018, <<https://www.rfc-editor.org/info/rfc8300>>.
- [RFC8402] Filsfils, C., Ed., Previdi, S., Ed., Ginsberg, L., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing Architecture", [RFC 8402](#), DOI 10.17487/RFC8402, July 2018, <<https://www.rfc-editor.org/info/rfc8402>>.

- [RFC8660] Bashandy, A., Ed., Filsfils, C., Ed., Previdi, S., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing with the MPLS Data Plane", [RFC 8660](#), DOI 10.17487/RFC8660, December 2019, <<https://www.rfc-editor.org/info/rfc8660>>.
- [RFC8754] Filsfils, C., Ed., Dukes, D., Ed., Previdi, S., Leddy, J., Matsushima, S., and D. Voyer, "IPv6 Segment Routing Header (SRH)", [RFC 8754](#), DOI 10.17487/RFC8754, March 2020, <<https://www.rfc-editor.org/info/rfc8754>>.
- [RFC8986] Filsfils, C., Ed., Camarillo, P., Ed., Leddy, J., Voyer, D., Matsushima, S., and Z. Li, "Segment Routing over IPv6 (SRv6) Network Programming", [RFC 8986](#), DOI 10.17487/RFC8986, February 2021, <<https://www.rfc-editor.org/info/rfc8986>>.

Authors' Addresses

Stewart Bryant (editor)
University of Surrey 5GIC

Email: sb@stewartbryant.com

Uma Chunduri (editor)
Intel Corporation

Email: umac.ietf@gmail.com

Alexander Clemm
Futurewei

Email: ludwig@clemm.org

